Modelling Energy Spot Prices: Empirical evidence from NYMEX

Nikos Nomikos [*] and Kostas Andriosopoulos [+]

ABSTRACT

This paper investigates the behaviour of spot prices in eight energy markets that trade futures contracts on NYMEX. We consider two types of models, a mean reverting model, and a spike model with mean reversion that incorporates two different speeds of mean reversion; one for the fast mean-reverting behaviour of prices after a jump occurs, and another for the slower mean reversion rate of the diffusive part of the model. We also extend these models to incorporate time-varying volatility in their specification, modelled as a GARCH and an EGARCH process. We compare the relative goodness of fit of the different modelling variations both in sample, using Monte Carlo simulations, as well as out-of-sample, in a Value-at-Risk (VaR) setting. Our results indicate the presence of a "leverage effect" for WTI, Heating Oil and Heating Oil - WTI crack spread, whereas for the remaining energy markets we find the presence of an "inverse leverage" effect. Also, the addition of the EGARCH specification for the volatility process improves both the in-sample fit as well as the out-of-sample VaR performance for most energy markets that we examine.

JEL Classification: C10; C51; C52; C63; Q40.

Key words: Energy Markets, Volatility Modelling, Mean Reversion Jump Diffusion, Leverage/ Inverse Leverage Effect.

[*] **Corresponding Author:** Nikos Nomikos, Cass Business School, City University, 106 Bunhill London EC1Y 8TZ, UK, Tel: +44 (0) 2070400104, Fax: +44 (0) 2070408681, e-mail: n.nomikos@city.ac.uk.

[+] Kostas Andriosopoulos, Faculty of Finance, ESCP Europe, 527 Finchley Road, London NW3 7BG; email: kandriosopoulos@escpeurope.eu

Acknowledgement: We would like to thank two anonymous referees for their helpful comments.

1. Introduction

Over the past decade significant changes have taken place in the world's energy markets. Changing economic patterns, globalization, international politics, war, technological advances and structural changes within the world's energy industry, have resulted in a volatile market environment which also increased the need of market participants for risk management using derivative contracts such as futures and options. In this volatile market environment, it is important for market participants to use risk management models that can capture the most significant risks in the market. However, due to the unique features of energy markets, the traditional approaches for modeling prices that are used in financial markets are not applicable. For instance, energy prices exhibit extreme movements and volatility over short periods of time and may also be characterized by spikes which occur due to short-term supply or demand shocks. In addition, energy prices have the tendency to mean-revert to a long-run equilibrium level. Given these stylized facts, the assumption used in the Black-Scholes-Merton model (Black and Scholes, 1973; and Merton, 1973) that the underlying asset follows a log-normal random walk may not be appropriate.

The mean-reverting process has been considered by many academics and practitioners as the natural choice for commodities. The reason is that, according to microeconomic theory, in the long run a commodity's price should be tied to its long-run marginal production cost; that is it tends to revert back to a "normal" long-term equilibrium level. There is a wealth of papers in the literature that confirm mean reversion in spot oil prices based on strong empirical evidence, such as Gibson and Schwartz (1990), Brennan (1991), Cortazar and Schwartz (1994) and Schwartz (1997). Evidence of mean reversion for energy and agricultural commodities comes also from the futures markets, e.g. Bessembinder et al (1995), Baker et al (1998), and Pindyck (1999). In addition, the analysis of volatility of asset prices is a research area that has been widely examined over the years by numerous studies, unveiling a number of stylized facts. According to Engle and Patton (2001), a good volatility model should be able to capture the most important stylized facts of an asset's volatility, which are mean reversion, volatility clustering, and persistence, the latter measured by calculating the volatility's half-life. Intuitively, we would expect to find that the innovations of the log-price series for all energy markets exhibit volatility clustering, and also that they have an asymmetric impact on the price volatility, with this asymmetry attributed to a leverage or risk premium effect.

In their study, Baumeister and Peersman (2008), when examining crude oil prices found that positive shocks, due to shifts in global demand, have greater impact on price volatility compared to negative shocks, which can be attributed to supply disruptions. This observation is consistent with the presence of an “inverse leverage” effect (Geman, 2005), which is also evident in the natural gas prices examined by Kanamura (2009), and in hourly electricity prices from Northern California examined by Knittel and Roberts (2005) using an EGARCH (1,1) model. Eydeland and Wolyniec (2003) in their study on a number of energy markets, also conclude that an “inverse leverage” effect should be expected. Hence, in the case of the energy markets we examine, it is expected that positive price shocks will have a greater impact on volatility than negative ones. Identifying any asymmetric tendencies in the volatility of the energy markets under investigation, using the EGARCH specification, can result in more efficient risk management applications by market practitioners and may also enhance the accuracy of various widely used risk management techniques, such as Value-at-Risk (VaR). Since volatility is an unobservable market variable, it is important to get the most accurate estimate in order to optimize our risk management models and eventually determine the best possible hedging strategies.

Considering the above, the motivation for this research mainly stems from the existing controversies in the empirical literature, as to which modeling approach is best for describing the behaviour of energy spot prices and capturing their risk characteristics. As a sound understanding of the stochastic dynamics of energy prices is a prerequisite for making an investment into energy commodities, we carry out a thorough empirical analysis by examining the performance, in terms of explanatory power and goodness of fit, of models that incorporate mean-reversion and spikes in the stochastic behaviour of the underlying asset. We consider two types of models: a mean reverting model, where prices have the tendency to revert to their long-run mean, and a spike model that incorporates two different speeds of mean reversion to capture the fast mean-reverting behaviour of returns after a jump occurs and the slower mean reversion rate of the diffusive part of the model. The different mean reversion rate is applied for a period of time equal to the half-life of jump returns for each energy market respectively. We also extend these models to incorporate time-varying volatility in their specification, modelled as a GARCH and an EGARCH process.

This paper contributes to the existing literature on modeling energy prices (see among others, Dixit and Pindyck, 1994; Schwartz, 1997; Clewlow and Strickland, 2000; Lucia and Schwartz, 2002; Cartea and Figueroa, 2005; Geman and Roncoroni, 2006; Cartea and Villaplana, 2008, Askari and Krichene, 2008) by expanding the choice of available models and the number of energy markets that these models are applied on. We use spot prices of the eight most traded energy futures contracts on NYMEX, covering the crude oil and all its by-product fuel markets, the soaring - due to their increased environmental importance - natural gas and propane markets, and one of the most liquid electricity markets. The performance of each model is assessed on the basis of how well it can capture the trajectorial and distributional properties of the real market process. In order to compare the aforementioned processes and identify which one describes the data best, we run Monte Carlo simulations to replicate the price paths, and then test the goodness of fit of the models using a variety of both quantitative and qualitative tests. In addition, the proposed models are evaluated out-of-sample in terms of their Value-at-Risk performance, using a two stage evaluation process. Moreover, we contribute in the existing literature by providing detailed information on the jump detection process, formally testing for any clustering and seasonality effects in the occurrence of jumps for all eight energy markets. This way, we provide a better understanding of how energy markets behave, what is the best modelling approach for each individual spot market and, consequently, the best model for the pricing of the relevant futures and options contracts. Identifying the correct dynamics for the energy prices is of great relevance for hedging, forecasting, and policy making in the energy markets. A further contribution in the literature is that we empirically test which model can sufficiently capture and describe the dynamics of the two 1-1 crack spreads of crude oil with fuel oil and gasoline that trade futures contracts on NYMEX. From the perspective of a petroleum refiner who operates between the crude oil and the refined products markets, modelling accurately the dynamic behaviour of the two crack spreads and their constituents is of utmost importance, since unexpected changes in the prices of the crude oil or the refined products can significantly narrow the spread and put refiners at enormous risk.

The structure of this paper is as follows. The next section presents the methodology used for modelling the spot energy markets under investigation and estimating the parameters for calibrating the models to real market prices. In section 3, the data and their properties are described. Section 4 offers empirical results, while section 5 evaluates the performance of each model in terms of matching the actual spot price behaviour. Section 6 presents the performance of the models in a VaR setting and finally, section 7 concludes this study.

2. Mean-Reverting Jump Diffusion GARCH/EGARCH Model

As already established, mean reversion is a main feature of energy commodities' event behaviour. In addition, energy prices often exhibit unexpected and discontinuous changes, so it is more appropriate to combine mean reversion and jump diffusion into the same model. The inclusion of spikes in the model is also justified by the existence of fat tails in the daily energy prices which suggests that the probability of rare events is much higher than the one implied by a Gaussian distribution; see for instance Cartea and Figueroa (2005) for a discussion on this in the UK power markets. According to the empirical findings presented in the literature, the presence of both excess skewness and kurtosis in all energy price returns suggests that a jump-diffusion model is more appropriate for both derivatives valuation (e.g. options pricing) and risk management purposes (e.g. VaR applications). Askari and Krichene (2008) point out that when jumps are added to oil price returns in a diffusion-based stochastic volatility model, sufficient variability and asymmetry in the short-term returns can be generated to match the skewness of implied volatility from short-term options. In their model, Clewlow and Strickland (2000) use the same speed of mean reversion for both spikes and normal shocks, inducing some persistence in the jumps especially when the mean-reverting coefficient is small. However, because the spikes represent a transitory phenomenon, after a jump has occurred prices do not stay at the high level to which they jump but tend to revert to their long-run mean. Consequently, when modelling energy prices it is also important to account for the fact that the decay rate of the jumps can be much faster than the decay rate of the diffusive component. We incorporate this feature in our model by using two different speeds of mean reversion, a fast one after a spike has occurred and a slower for the normal (diffusive) shocks.

Another issue that needs to be addressed in our modelling methodology is the behaviour of volatility, which exhibits high values and clustering. Cartea and Villaplana (2008) in all three electricity markets they examine find that prices follow a strong seasonal component and thus a model with seasonal or time-varying volatility is preferable than one with constant volatility. Thus, in accordance with the empirical evidence from various studies related to the energy markets, we use constant, as well as GARCH (Bollerslev, 1986) and EGARCH

(Nelson, 1991) specifications for the variance. Our mean-reversion jump diffusion model that incorporates the observed stylised facts of energy prices and their volatility is based on Schwartz's (1997) one-factor model. The model is extended to allow for a deterministic seasonality as in Lucia and Schwartz (2002) and Cartea and Figueroa (2005). We assume that log-prices can be expressed as the sum of a predictable and a stochastic component as follows:

$$\ln S_t = g(t) + Y_t \tag{1}$$

with the spot price represented as:

$$S_t = G(t) e^{Y_t} \tag{2}$$

where $G(t) \equiv e^{g(t)}$ is the predictable component of the spot price S_t that takes into account the deterministic regularities in the evolution of prices, namely seasonality and trend. Also, Y_t is a stochastic process whose dynamics are given by the following equation:

$$dY_t = a_i(\mu - Y_t)dt + \sigma_t dZ_t + kdq_t \tag{3}$$

where a_i is the mean reversion rate, μ is the long-term average value of $\ln S_t$ in the absence of jumps, σ_t is the volatility of the series, dZ_t is a Wiener process, k is the proportional jump size and dq_t is a Poisson process. It is assumed that the Wiener and the Poisson processes are independent and thus not correlated, which further implies that the jump process is independent of the mean-reverting process.

Using equations 1 and 3, we follow Dixit and Pindyck (1994) and after applying Ito's Lemma our model can be discretised in the following logarithmic form:

$$\ln S_t = g_t + \left(\ln S_{t-1} * e^{-a_i \Delta t}\right) + \left(\ln \bar{S} - \frac{\sigma_t^2}{2a_i}\right) * \left(1 - e^{-a_i \Delta t}\right) + \sigma_t * \sqrt{\frac{1 - e^{-2a_i \Delta t}}{2a_i}} * \varepsilon_1 + J(\mu_J, \sigma_J) * I_{(u_t < \Phi \Delta t)} \tag{4}$$

where,

$$a_i = \begin{cases} a_1 = a_{JD}, & \text{when a jump occurs; for a duration equal to jump returns' half-life} \\ a_2 = a, & \text{otherwise} \end{cases} \qquad i = 1, 2 \tag{4.1}$$

$$\sigma_t = \begin{cases} \sigma_t = \sigma \; [\text{Constant}] \\ \sigma_t = \sqrt{\beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 * \sigma_{t-1}^2} \; [GARCH(1,1)] \\ \sigma_t = \sqrt{e^{\beta_0 + \beta_1 * \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \beta_2 * \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_3 * \ln\left(\sigma_{t-1}^2\right)}} \; [EGARCH(1,1)] \end{cases} \tag{4.2}$$

$$g_t = \gamma_0 \sin\left(\frac{2\pi(t+\tau)}{252}\right) + \gamma_1 t \tag{4.3}$$

$$I_{(u_t < \Phi\Delta t)} = \begin{cases} 1 \text{ when } u_t < \Phi\Delta t, \text{ i.e when a jump occurs} \\ 0 \text{ when } u_t > \Phi\Delta t, \text{ i.e when there is no jump} \end{cases}$$

$$\text{J} \sim N(\mu_J, \sigma_J) \text{ with Mean}: \mu_J = (\bar{\kappa}_J + \sigma_J \varepsilon_2) \text{ and Standard Deviation}: \sigma_J \tag{4.4}$$
$$\varepsilon_1, \varepsilon_2 \sim N(0,1), \text{ and } u \sim U[0,1]$$

where $\ln \bar{S}$ is the long-run mean (μ), Φ is the average number of jumps per day (daily jump frequency), $\bar{\kappa}_J$ is the mean jump size, σ_J is the jump volatility, ε_1 and ε_2 are two independent standard normal random variables, and u is a uniform [0, 1] random variable. The term $\text{I}_{(u_t < \Phi\Delta t)}$ is an indicator function which takes the value of 1 if the condition is true, and 0 otherwise. This condition leads to the generation of random direction jumps at the correct average frequency. When the randomly generated number is below or equal to the historical average jump frequency, the model simulates a jump with a random direction; no jump is generated when the number is above that frequency. When a jump occurs its size is the mean size of the historical jump returns plus a normally distributed random amount with standard deviation σ_J. Notice as well that our modelling approach allows for the possibility of both, positive and negative jumps to occur[1].

In addition, our model takes into account the fact that most energy prices exhibit a seasonal behaviour that follows an annual cycle. Various methods have been used in the literature for the deterministic seasonal component, from a simple sinusoidal (Pilipovic, 1998) or a constant piece-wise function (Pindyck, 1999; Knittel and Roberts, 2005), to a hybrid of both functions (Lucia and Schwartz, 2002; Bierbrauer et al., 2007). We account for this periodic behaviour

[1] Merton (1976) in his original jump diffusion model assumes that the jump size distribution is lognormal, and so jumps can occur in only one direction (positive jumps).

by fitting a sinusoidal function with a linear trend to the actual prices, as described by g_t. The estimation is done using Maximum Likelihood (ML), with the sine term capturing the main annual cycle, and the time trend capturing the long-run growth in prices[2]. Moreover, we incorporate in our model the possibility for the returns to have a different mean reversion rate after a jump occurs. This approach is in line with Nomikos and Soldatos (2008) who use two different coefficients of mean reversion, one for the normal small shocks and another, larger, for the spikes to capture the fast decay rate of jumps observed in the energy markets. Geman and Roncoroni (2006) also analyse the existence of different speeds of mean reversion in the context of mean-reverting jump-diffusion models for three major US power markets, by introducing a class of discontinuous processes exhibiting a "jump-reversion" component to represent the sharp upward moves that are shortly followed by drops of the same magnitude. Our approach is flexible enough to accommodate the fact that the abnormal events that cause the jumps have different effect in each market and hence, prices tend to remain at the level to which they jump for a longer or shorter period of time, depending on the energy market under investigation. Therefore, prices following a jump are adjusted by using in equation (4) a different mean reversion rate, noted as a_{JD}, for a period of time equal to the half-life of jump returns for each energy market; when another jump occurs within the duration of the half-life period used, then a_{JD} is used again for the same number of days, counting from the day following the last jump (see equation 4.1). If no other jump occurs within that period, then a_2 is used until a new jump occurs. Incorporating the half-life measure in this way, allows for the model to better reflect the duration of both short- and long-term shocks of different magnitudes, exhibited in energy prices. This results in a more flexible framework, compared to the model proposed by Nomikos and Soldatos (2008) which fits best mainly the highly volatile electricity markets, as the speed of mean reversion estimated after a spike shock is significantly higher than the normal mean reversion rate. In addition, the model we propose incorporates in its specification GARCH and EGARCH volatility, to account for volatility clustering and any asymmetries that are usually observed in energy prices.

Regarding the mean-reverting part of equation 4, we use an exact discretization for the simulations since the presence of jumps complicates the use of a large Δt. This is because the

[2] We follow the approach used in Pilipovic (1998) to calculate the seasonal component in the data, because this method is more flexible than using dummy variables. According to Lucia and Schwartz (2002) the use of dummy variables does not provide a smooth function for the seasonal component observed in the data, which can cause discontinuities when pricing forward and futures contracts.

drift of the mean-reverting process is a function of the current value of a random variable and in order to simulate the jumps correctly the time step Δt must be small relative to the jump frequency. Because we want to model rare large jumps, if the time interval Δt is sufficiently small, the probability of two jumps occurring is negligible $\left((\phi\Delta t)^2 \ll \phi\Delta t\right)$. That makes it valid to assume that there can be only one jump for each time interval; in our case one every day since Δt is equal to one day. Especially when Δt is increased to one week or one month, as it is usually the case with real option applications that involve pricing medium- and long-term options, it is more important to use an exact discretization for the simulation process, because the overall error from the first-order Euler and the Milstein approximations will be much higher [3]. The random number generation of the Monte Carlo (MC) simulations already introduces an error in our results, therefore using these approximations that need a very small Δt and thus also introduce a discretization error, would lead to higher computational cost into the simulations.

As for the two time-varying volatility model specifications of equation 4.2, in the case of the GARCH process, ε_{t-1}^2 represents the previous periods' return innovations and σ_{t-1}^2 is the last period's forecast variance (GARCH term). As for the EGARCH process, β_0 denotes the mean of the volatility equation. The coefficients β_1 and β_2 measure the response of conditional volatility to the magnitude and the sign of the lagged standardised return innovations, respectively; as such, these coefficients measure the asymmetric response of the conditional variance to the lagged return innovations. When $\beta_2 = 0$, there is no asymmetric effect of the past shocks on the current variance, while when $\beta_2 \neq 0$ asymmetric effects are present in response to a shock; for instance, $\beta_2 > 0$ indicates the presence of an "inverse leverage" effect. Finally, β_3 measures the degree of volatility persistence. Knittel and Roberts (2005) suggest that a positive shock in electricity prices represents an unexpected demand shock which has a greater impact on prices relative to a negative shock of the same size, as a result of convex marginal costs and the competitive nature of the market. Moreover, Kanamura (2009) suggests that this inverse leverage effect, i.e. positive correlation between prices and volatility, is a phenomenon often observed in energy markets, whereas evidence from the

[3] Clewlow & Strickland (2000) use the first-order Euler's approximation in order to get the discrete time version of the Arithmetic Ornstein-Uhlenbeck: $x_t = x_{t-1} + a*(\bar{x} - x_{t-1})*\Delta t + \sigma*\sqrt{\Delta t}*\varepsilon_t$ where the discretization is only correct in the limit of the time step tends to zero.

stock markets suggests that the opposite relationship exists between volatility and prices, namely the "leverage" effect[4]. Hence, intuitively we expect the asymmetry parameter to be positive and significant for most energy markets, implying that positive shocks have greater effect on the variance of the log-returns compared to negative shocks, consistent with the presence of an "inverse leverage" effect.

Finally, the different models used for modelling the spot prices of the energy markets are summarized in Table 1; "GBM" stands for Geometric Brownian Motion; "MR" for Mean Reversion; "MRJD" for Mean Reversion Jump Diffusion; "OLS" for Ordinary Least Squares (constant volatility).

3. Description and Properties of the Data

Before discussing the estimation results for our various modeling specifications, let us look at the data to verify whether the stylized facts that we aim at reproducing are indeed present. We investigate the behaviour of the spot prices of eight of the most important energy markets that trade futures contracts on NYMEX, each one of them having its unique impact on the worldwide marketed energy supply and demand. We collect spot daily prices from Thomson DataStream, which are the official closing prices of the 1st nearby futures contract issued by the NYMEX, for the period 12/09/2000 to 01/02/2010 for the following contracts:

1. Heating Oil, New York Harbour No.2 Fuel Oil, quoted in US Dollar Cents/Gallon (US C/Gal); hereafter named as "HO";
2. Crude Oil, West Texas Intermediate (WTI) Spot Cushing, quoted in US Dollars/Barrel (US$/BBL); hereafter named as "WTI";
3. Gasoline, New York Harbour Reformulated Blend stock for Oxygen Blending (RBOB), quoted in US C/Gal; hereafter named as "Gasoline";
4. 1-1 Crack Spread of Gasoline with WTI, quoted in US $/BBL; hereafter named as "CS_Gasoline_WTI"[5];

[4] The "leverage effect" terminology is first used by Black (1976) who suggests that negative shocks on stock prices increase volatility more than positive ones. The intuition behind it is that a lower stock price reduces the value of equity relative to debt, thereby increasing the leverage of the firm and thus making it a more risky investment.

[5] The spot series of the two 1-1 crack spreads with the WTI have been constructed after converting the Fuel Oil and Gasoline spot prices that are quoted in US C/gallon into US $/Barrel, taking into account that there are 42 gallons in one barrel and 100 cents per dollar. Then, the two series are rebased to 100 so they can later be transformed to logarithmic prices and apply our modelling methodology.

5. 1-1 Crack Spread of Fuel Oil with WTI, quoted in US $/BBL; hereafter named as "CS_HO_WTI";
6. Natural Gas, Henry Hub, quoted in US Dollars/Million British Thermal Units (US$/MMBTU); hereafter named as "NG";
7. Propane, Mont Belvieu Texas, quoted in US C/Gal; hereafter named as "Propane";
8. PJM, Interconnection Electricity Firm On Peak Price Index, quoted in US Dollars/Megawatt hour (US $/Mwh); hereafter named as "PJM".

Form the total sample, the period 12/09/2000 to 12/09/2007, consisting of 1,827 observations, is used as the in-sample period, whereas the period 13/09/2007 to 1/02/2010, consisting of 623 observations, is used as the out-of-sample testing period. The remaining analysis on the properties and descriptive statistics of the data relates only to the in-sample period while the VaR performance is assessed in the out-of-sample period. Figure 1 shows the time series of the spot log-prices and their first differences. We see that all energy markets are very volatile and some of them, such as the PJM, Heating Oil crack spread, Natural Gas and the Propane markets, seem to exhibit more distinctive jumps in their price behaviour. Moreover, the graphs indicate a distinct upward trend, which is more obvious for the WTI, Gasoline, and Heating oil markets, reflecting the continuous rally in commodity prices during the second part of our sample. A rigid supply, in combination with an expanding global demand for crude oil and its by-products resulted in big demand-supply imbalances, which in turn led to the great variability observed in energy prices. Finally, when looking at the spot log-price differences we see that most of the series vary with time and also form clusters, which indicate the presence of time-varying volatility.

Descriptive statistics are estimated for the natural logarithm of the spot prices and reported in Table 2 for both the spot price series in logarithmic levels (Panel A) and their first differences (Panel B). As can be seen in panel B, the annualized volatility (as measured by the standard deviation of log-returns) of most energy markets ranges from 13% for the Heating Oil – WTI crack spread to 240% for PJM, which is significantly larger than the typical volatility observed in financial markets. Overall, the two crack spreads have lower volatility than the outright series due to the high correlation between the prices of their constituent contracts. Looking at panel A of Table 2, is observed that for all energy markets, with the exception of NG and Propane, the skewness is positive, indicating that extreme high values are more probable than low ones. Turning next to the log-price changes, the results regarding the coefficients of skewness are mixed, with all three fuel markets and the crack spread of WTI with Gasoline being negatively skewed, whereas the rest of the energy markets are positively skewed (see panel B, Table 2). Also, the coefficient of kurtosis, which gives an indication of the probability of extreme values, is above three for all energy markets, implying that log-returns are leptokurtic; this suggests that the probability of extremely high or low returns is much higher than that assumed by the normal distribution. This effect is more obvious for the NG, Propane, PJM and the two crack spreads in which case the high value of the coefficient of kurtosis (between 10.67 and 45.15) is indicative of spikes in the price series.

As a result, the assumption of normality is overwhelmingly rejected for all the energy markets, on the basis of the Jarque-Bera (1980) test which is significant at the 1% level. It is obvious that non-normality occurs mostly due to the large price movements and spikes in all energy markets that eventually lead to fat tails. Moreover, looking at panel A in Table 2, we see that the average logarithmic price for most energy markets is reduced when the filtered series is examined (i.e. when jumps are excluded) indicating that jumps have, on average, a positive impact on log-prices [6]. The only exceptions are the WTI and Gasoline markets where jumps have a negative impact on log-prices. In panel B we also report the Ljung-Box (1978) $Q(k)$-statistic and Engle's (1982) ARCH test ($Q^2(k)$-statistic), where we test the significance of autocorrelation in the returns and squared returns for lags one and 20, respectively. From the reported values there is evidence of serial correlation for most of the log-return series with the exception of WTI and Gasoline. In addition, based on Engle's ARCH test we find significant serial correlation in the squared log-returns of all energy markets, which indicates the presence of time-varying volatility in the return series.

Finally, in order to identify whether the series are mean reverting, a comparison procedure known as "confirmatory data analysis" is performed, where two tests for unit root non-stationarity, the Augmented Dickey-Fuller (ADF; Dickey and Fuller, 1979) and the Philips-Perron (PP; Phillips and Perron, 1988), and one test for stationarity, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS; Kwiatkowski et. al, 1992), are employed. For the results to be robust, all three tests should give the same conclusion. From the results in panel A of Table 2 we can infer that the price-levels of most energy markets are not stationary, a conclusion confirmed by all three tests; the only exceptions are the two crack-spreads and the PJM

[6] A detailed discussion on how the filtered series is estimated is given in the following section.

markets where price levels appear to be stationary on the basis of the ADF and PP tests. On the other hand, in Panel B of Table 2 we can see that the first differences of the spot log-price series are strongly stationary for all energy markets, indicating the presence of mean reversion in the series. This conclusion, although it may not have been expected due to the presence of jumps in most of the energy series, can be justified by the fact that these jumps seem not to affect the stationarity of the series because they are short-lived and price levels eventually revert to their mean after a jump has occurred.

4. Empirical findings

The input parameters for the Monte Carlo simulations are estimated from the historical spot price series of the different commodities. We consider first the jump parameters. Estimating the jump parameters, especially for energy prices, can be quite complicated because usually there is no indication of the exact time the jump will occur, and thus jumps can only be observed as part of the historical spot time series. There are two widely used approaches for estimating the jump parameters, the first being the Recursive Filter (R-F) (Clewlow and Strickland, 2000; Clewlow et al 2000), and the second being the Maximum Likelihood (M-L) (Ball and Torous, 1983). Empirical analysis suggests that the R-F estimation method can be superior to the M-L method when it comes to estimating jump parameters in energy markets; this is because the former method can pick the lower frequency, higher volatility jump components, instead of the higher frequency, lower volatility jumps that are estimated better with the latter. According to Clewlow and Strickland (2000), a potentially undesirable property of the M-L method is that it tends to converge on the smallest and most frequent jump components of the actual data. As energy price return series exhibit jumps that range from very high frequency and low volatility to low frequency and high volatility, it is important to be able to efficiently capture the latter ones.

Therefore, given that jumps in the energy markets are relatively infrequent but of large magnitude, the R-F method is more appropriate. Correct identification and measurement of jumps is very important. For instance, Nomikos and Soldatos (2008) point out the importance of spikes in electricity prices especially for market suppliers because, although their costs depend on the variable price for electricity, their revenues are mainly fixed; in fact, these rare spikes are the most important motive for hedging in the energy markets. In addition, these rare but large returns, significantly affect the value of medium- and long-term energy real investments, as is the case for example when pricing an undeveloped oil field. In particular, according to Dias (2003), the two main sources of uncertainty in an oilfield development project are fluctuations in the oil prices (market uncertainty), and variations in the volume and quality of the reserves (technical uncertainty). A mean-reverting model with jumps can capture both the mean-reverting price evolution of the underlying resources, as well as the sudden changes in prices due to unexpected news in the market.

The R-F algorithm is then implemented as follows: By assuming that jumps are relatively infrequent and that the diffusive volatility can be estimated based on the sample standard deviation of returns, we identify as jumps those “extreme” returns that are more than three standard deviations away from the mean, consistent with most studies in the literature. Now, given that we have identified some of the returns as jumps, we calculate a new estimate of the diffusive volatility by recalculating the sample standard deviation of returns, after filtering out those returns previously identified as jumps. During the filtering process, when a jump is identified, its respective log-price is being removed from the series and then replaced by the average of the previous and the next log-price. Then the new returns are calculated based on the filtered series. The new calculation gives us a lower estimate of the diffusive volatility and, based on that lower volatility, we repeat the same procedure in order to identify new jump returns. The process is repeated until the estimates converge and no further jumps can be identified. Finally, we calculate the jump parameters necessary for calibrating our models, on an annual basis, from the following relationships:

ϕ = Number of jump returns/ Time period of the data

$\bar{\kappa}_J$ = Average jump size of returns

σ_J = Standard deviation of jump returns

Panel A of Table 3 presents the estimated jump parameters used in the MRJD models, as calculated by the Recursive Filter algorithm; these parameters include the jumps’ daily frequency (Φ), daily standard deviation (σ_J) and average jump size ($\bar{\kappa}_J$). We observe that the average size of the jump returns is negative for the WTI, Gasoline, and PJM markets, whereas for the rest is positive. As for the daily jump frequencies, the highest frequency is observed for the crack spread of WTI with Gasoline, followed by the other volatile markets, i.e. the gas and electricity markets. Finally, in terms of the jumps’ volatility we see that the highest daily standard deviation values are calculated for the Gasoline (10.78%), Natural Gas (16.14%) and PJM (51.94%) markets, which are also the markets with the highest unconditional volatilities as evidenced in Table 2.

In addition, we test whether jumps arrive at random intervals, like the model predicts, or come in clusters, by comparing the arrival rate of daily jumps, as identified by the R-F methodology, to the arrival rate of a random series of jumps generated by a Poisson process with a frequency equal to the frequency of jump occurrence as reported in panel A of Table 3 for each energy market. The Kolmogorov-Smirnov (K-S) test statistics for the null hypothesis that the two samples are from the same distribution are reported in Table 4, where we can clearly see that the null hypothesis cannot be rejected for any of the energy markets. This confirms that there is no clustering behaviour observed in the occurrence of jumps for all markets examined. Similarly, we also test whether there is a seasonality in the occurrence of jumps; for instance it may be the case that jumps in certain markets, such as natural gas, may be more frequent in the winter months than in the summer months. For that we regress the number of jumps in each quarter against quarterly dummies, for the seven year period examined, and for all energy markets. None of the energy markets is found to exhibit any seasonality during each of the four quarters.[7].

Finally, for comparison purposes, the jump parameters were also calculated using the Maximum Likelihood Estimation method, based on the methodology by Ball and Torous (1983) and Weron and Misiorek (2008). On average, the volatility of jumps identified by the M-L method is smaller, consistent with the intuition that the R-F method is able to capture the larger in size jumps; in addition, the average jump size detected by the M-L method is smaller than the average jump size detected with the R-F method and the frequency of jumps detected with the M-L method is larger than that estimated with the R-F method. These results are consistent with the tendency of the M-L method to converge on the smallest and most frequent jump components of the actual data. Since through our modeling procedure we want to capture the low frequency but high volatility jumps in energy markets we use the jumps identified through the R-F procedure in the ensuing analysis[8].

[7] Additionally, six-month jumps' data, representing the cold and warm seasons, were also regressed against seasonal dummies, with the results confirming again that there is no seasonality effect in the occurrence of jumps for most of the energy markets.
[8] Results for these tests are available from the authors.

Turning next to the coefficients of mean reversion, these are estimated using a modified version of equation (3), following the methodology used by Dixit and Pindyck (1994):

$$\Delta x_t = a_0 + a_1 x_{t-1} + \varepsilon_t \ ; \qquad \varepsilon_t \sim N(0, \sigma_{regres.}) \tag{5}$$

where $x_t = \ln S_t$. Because we want to estimate the diffusive risk of the model, the regression is applied to the filtered (i.e. without jumps) series when considering the MRJD models; the filtered series is the returns series that excludes all returns that have previously been identified as jumps. In the case of the simple MR models, the regression of equation 5 is applied to the un-filtered (i.e. with jumps) series. Then, for both cases, we calculate the estimates for a and σ using the following equations:

$$a = -\ln(1 + \hat{a}_1) \tag{6}$$

$$\sigma = \sigma_{regres.} \sqrt{\frac{2\ln(1+\hat{a}_1)}{(1+\hat{a}_1)^2 - 1}} \tag{7}$$

The long-term mean (μ) is calculated from the un-filtered historical time series of each commodity for all models. In order to estimate the mean reversion rate used after a jump occurs, we estimate the following regression on the un-filtered series:

$$\Delta x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-1} DUM_t + a_3 TIME_t + \varepsilon_t \ ; \qquad \varepsilon_t \sim N(0, \sigma_{regres.}) \tag{8}$$

where DUM_t is a dummy variable that takes the value of one when a jump occurs and zero otherwise, irrespective of the jumps' direction. We include a linear time trend in the regressions to allow for gradual shifts in the "normal" price (Pindyck, 1999)[9]. The trend coefficient is significant, albeit small in size, in all cases except for the two crack spreads. The presence of a trend in those series is also confirmed visually by looking at the graphs in Figure 1. Therefore, we use the de-trended series to estimate the different speeds of mean reversion and capture the real expected evolution of the log-price series. The mean reversion

[9] We have also used in our regressions the quadratic trend model, which is another extrapolation model commonly used for commodities, however the regression coefficients of the additional term t^2 were insignificant for all the energy markets considered in our study.

rate after a jump occurs is then calculated from the coefficients of equation (8) using the following formula:

$$a_{JD} = -\ln\left(1 + a_1 + a_2\right) \tag{9}$$

All estimates are annualized assuming 252 trading days per year. Finally, one important parameter of the mean reverting process is the half-life, defined as the time required for the log-price to go back half way to its long-run mean from its current level, subject to no other shocks occurring, and is estimated using the following equation:

$$t_{\frac{1}{2}} = \frac{\ln(2)}{a_i} \tag{10}$$

$$a_i = \begin{cases} a_1 = a_{JD}, \text{ for returns identified as jumps} \\ a_2 = a, \text{ for smooth returns} \end{cases} \qquad i = 1, 2$$

Panel B of Table 3 presents the two mean reversion rates and the daily standard deviations used in the MR and MRJD models, for all energy markets. A general observation is that the estimated mean reversion rate for the returns following a jump is higher for all markets, compared to the diffusive mean reversion rate, which indicates that when a jump occurs prices tend to revert back to their long-term mean faster. The high speed of mean reversion for the spikes is one of the significant features of this model, which also improves the fit of the model to the observed prices in the market. In addition, the estimated mean reversion rate for the un-filtered series is higher when compared to that of the filtered series, suggesting that when spikes are extracted from the sample the coefficient of mean reversion decreases. The exception to that are the three fuel markets (WTI, Heating Oil and Gasoline) and Propane, where the daily mean reversion rate estimated for both the un-filtered and filtered series is similarly small for all three, in the range of 0.1% to 0.2%. This observation reflects the fact that for the seven year period examined, the fuel markets exhibit a distinctive upward trend, with a small tendency to revert to a long-term mean. However, when looking at the α_{JD} values these are in the range of 0.8% for Propane (the smallest rate amongst the eight energy markets) to 2.1% for Gasoline, indicating that after a jump occurs prices do tend to revert faster to their long-term mean.

We also note that the highest speed of mean reversion for both the un-filtered and filtered series occurs for the PJM market, which is also the most volatile market with estimated daily volatility of 15.8% and 13.2%, respectively. When we compare the speed of mean reversion for the spikes amongst the eight energy markets, we see that PJM has the highest (11.5%), followed by the Heating Oil - WTI crack spread (4.1%). This means that following a positive (negative) jump, prices will be reduced (increased) by 11.5% and 4.1%, respectively each day in order to return to their long-term mean. However, when the impact of the spikes has died-out, prices will revert to their mean at a much lower daily rate of 5.5% and 1.3%, respectively. This is consistent with the stylised fact of energy markets that, following a jump, prices quickly revert back to their long-run mean at a faster rate than when a normal shock occurs.

The results for the calculated half-lives, in days, of the smooth and jumpy returns are also presented in panel B. The half-lives of the jumpy returns are calculated using equation 10 and represent the respective durations we are using in our MRJD models for the higher mean reversion rate (α_{JD}) after a jump occurs. We can see that for all energy markets the half-lives of the jumpy returns are much shorter than the ones for the smooth returns; also, the smallest half-life duration for the jumpy returns is observed for the PJM market (6 days), followed by the crack spread of Heating Oil - WTI (17 days), reflecting the higher mean reversion rates observed in those markets. This is expected as the PJM is the most volatile market which experiences frequent and sudden positive and negative jumps, bringing smooth returns back to their long-term level faster, when compared to the other energy markets. The highest half-life duration of jumps is that of Propane (87 days) followed by NG (72 days). For the fuel markets, the half-life of the jumpy returns for WTI, HO, GASOLINE and the Gasoline – WTI crack spread is 36, 67, 34 and 26 days, respectively. Finally, we also note that, as expected, when jumps are removed from the series the estimated volatility is reduced for all energy markets which means that spikes play a very significant role in terms of explaining the volatility in the market.

Turning next to the volatility estimates, the coefficient estimates for the GARCH(1,1) and EGARCH(1,1) models, using equation 5 for the specification of mean, are presented in Table 5. The regression is applied to both the un-filtered and filtered series, with the estimates used for the MR and MRJD models, respectively. Because results are qualitatively similar, only those estimated from the un-filtered series are reported in the table. All GARCH coefficients are significant at the 5% level, verifying the presence of time-varying volatility in all energy markets. In addition, we observe that the sum of the coefficients β_1 and β_2 for the GARCH models is greater than the coefficient β_3 of the EGARCH model, indicating that the volatility persistence in the latter case is reduced, which is consistent with the literature on volatility models. Looking at the estimates for the β_2 coefficients of the EGARCH models, which measure the leverage effect, we can see that they are significant in all cases indicating the presence of asymmetries in the way past shocks affect the current volatility. For the WTI, Heating Oil and Heating Oil – WTI crack spread returns, the coefficient estimate β_2 is negative at the five percent level, indicating the presence of a "leverage" effect; in other words negative shocks have greater impact on volatility than positive shocks. One possible explanation for this finding may be that price shocks for the aforementioned markets are more supply- than demand-driven, due to the fact that the market has been operating at the steep part of the supply stack in recent years. This phenomenon can be attributed to the very low spare capacity in world energy production, with small supply disruptions causing large price increases due to difficulties of rapid replacement of any production shortfalls. This is in contrast to what one expects to find in commodity markets as well as recent empirical evidence by among others Baumeister and Peersman (2008) who point out that oil price surges can almost entirely be explained by shifts in global demand (positive shocks), with the contribution of supply shocks (negative shocks) on crude oil price volatility diminishing considerably over the recent years. This inconsistency in the findings can be attributed to the fact that over the past few years other exogenous factors, in addition to the market fundamentals of supply and demand, have been driving the oil markets. As a result, the fuel markets in particular have become more prone to movements of a much broader range of financial indicators like international currencies' exchange rate movements relative to the US dollar, interest rates, equity markets' performance, as well as the widespread use of "paper" derivative products both for the purposes of risk management as well as for speculation.

For the remaining energy markets, the asymmetry parameter is positive at the 5% significance level, which implies that positive shocks, as described by unexpected demand shocks, have greater impact on volatility compared to negative shocks, which is consistent with the presence of an "inverse leverage" effect. We can argue that since the beginning of the new millennium, worldwide economic growth gave rise to stronger than expected demand for energy products that are critical to the global economy. As a result, demand outpaced the near-term ability of the market to bring forth proportionate additional supplies; the resulting tightness in the global energy markets caused prices to increase, and the impact of this increase has been felt throughout the whole chain of production. Along the same lines, Kanamura (2009) finds that demand for US natural gas prices is highly inelastic in the short-term, with the energy use being independent of the price change, suggesting the presence of an "inverse leverage" effect. So, when an unexpected demand shock occurs, energy prices are expected to exhibit this "inverse leverage" effect, a conclusion that can be drawn from our results; this is also consistent with the findings in Eydeland and Wolyniec (2003) regarding the energy markets.

5. Simulation of Estimated Models

After estimating the parameters of the model, we use Monte Carlo (MC) to simulate the behaviour of each market; the simulations are carried out based on equation 4 and the paths are simulated 100,000 times. The starting date of the simulations is the same as the initial date of our historical prices, i.e. 12/09/2000, with the horizon of the simulated distribution extending up to 12/09/2007; in total 1827 trading days. Since the main purpose of this paper is to propose models that can capture the distributional characteristics of the underlying market, MC simulation is a valuable tool for helping with the selection criteria of the best model. Clewlow et al. (2000; 2001) use Monte Carlo simulations on different variations of the MRJD model and demonstrate how these models can be used to price energy options whose payouts are path-dependent, or rely on multiple energies. In addition, other applications of MC simulation include pricing of various energy derivatives contracts, policy development and risk monitoring. Hence, because we want to determine whether our models can capture the major characteristics of the distribution of energy spot prices, in what follows we perform a distribution analysis which will help us analyze the price behaviour over a period of time and, at the same time, assist us with testing, benchmarking, and selecting the most appropriate model for describing each one of the energy markets we examine.

The descriptive statistics of the actual log-returns' series, along with the average per time-step simulated paths for all models used in our analysis, are presented in Table 6. The average of the simulated values at time t across all possible paths is calculated as:

$$S_t^s = \sum_{\omega=1}^{n} \frac{S_{t,\omega}^s}{n} \tag{11}$$

where, $S_{t,\omega}$ is the simulated spot price of path ω at time t, and n is the number of MC simulations. From Table 6 we see that for almost all the energy markets, the models that most closely match the skewness and kurtosis of the underlying distributions are the ones that incorporate jumps, namely the MRJD-OLS and the MRJD-EGARCH. It can also be noted that in the case of WTI, the skewness produced by the MRJD-OLS model is identical to the actual one, whereas the kurtosis value is the highest among the competing models, thus also following very closely the actual one. It is only in the Heating Oil and Propane markets that the MR-GARCH(1,1) model is able to better match the skewness and kurtosis of the actual price path. Therefore, it seems that our approach to allow for a different speed of mean reversion after a jump occurs, and to also extend the models to incorporate time-varying volatility in their specification modelled as an EGARCH process, improves the fit that our models have in terms of capturing the skewness and kurtosis of the actual series, for almost all energy markets.

To formally compare the actual returns' distribution with the average of the simulated series per time-step, we calculate the two-sample Kolmogorov-Smirnov (K-S) test. The two-sample K-S test is a non-parametric test for the equality of two probability distributions. The test effectively compares the distance between the actual and the simulated distribution around their mean, and the reported statistic is the maximum vertical deviation between the two curves. One of the advantages of the K-S test is that the value of the statistic is not affected by scale changes like using the logarithm of prices, as is the case in our data; it is a robust test that only considers the relative distributions of the data. In our case, the first sample $X_1,...,X_m$ of size $m=1826$ observations, which are the actual spot log-price returns, has a distribution with cumulative density function (c.d.f.) $F(x)$, and the second will be in every case the average per time-step simulated sample $Y_1,...,Y_m$ of the same size $m=1826$, having a distribution with c.d.f. $G(x)$. The null hypothesis of the K-S test is that F and G are from the same continuous distribution, with the alternative hypothesis that they are from different continuous distributions:

$H_0: F = G$ vs. $H_1: F \neq G$

Results from the K-S tests are also presented in Table 6; based on the calculated K-S test statistic we accept the null hypothesis that the actual and the average per time-step simulated distributions are identical at the 5% significance level, for the Gasoline, the two crack Spreads of WTI with Heating Oil and Gasoline, and the PJM markets. This is true for most models with the exception of the GBM where the null hypothesis of equality of distributions is overwhelmingly rejected. Comparing the values between the different models we can see that generally the models that incorporate jumps have the lowest value for the K-S test indicating that, at least nominally, these provide the closest match to the underlying distribution. For the remaining markets, although the null hypothesis that the samples are drawn from an identical distribution is rejected, the value of the K-S statistic is lower for the models that contain jumps in their terms. Furthermore, in Table 6, the models with the smallest K-S test-statistic value are indicated with a (+). It can be seen that the models producing the lowest K-S test-statistic are the MRJD-EGARCH(1,1) for WTI, Heating Oil, Heating Oil-WTI crack spread, Natural Gas, and Propane markets, the MRJD-OLS model for Gasoline and Gasoline-WTI crack spread, and finally the MR-EGARCH(1,1) for the PJM market. Overall, from the distributional comparison of the actual log-price returns and the average per time-step simulated returns, we can conclude that the addition of jumps in the simple mean reversion model - while allowing for a different speed of mean reversion after a jump occurs for a period of time equal to the estimated half-life of the jumpy returns - as well as the addition of the EGARCH (1,1) process, improves the fit of the simulated returns to the actual distributions, for most of the energy markets under investigation.

Furthermore, the relative goodness of fit for the various models is assessed by examining how closely each endogenous variable from our simulations tracks the actual spot logarithmic prices for the seven year period we examine. Clewlow and Strickland (2000) use the likelihood ratio test and the Schwartz Bayesian Information criterion to compare their various models. In our case, because we want to test the simulations' goodness of fit, we use three quantitative and one qualitative measure to check how closely the individual variables track their corresponding data series. The three quantitative measures are the root-mean-square error (RMSE), the root-mean-square percent error (RMSE %), and Theil's inequality coefficient (Theil's U) (Theil, 1961). The RMS error measures the deviation of the average simulated log-price from its actual time path, while the RMS percent error evaluates the magnitude of the RMS error as a percentage of the underlying spot price; finally, Theil's U measures the RMS error in relative terms.

Table 7 presents the comparison results for our models based on the RMSE, RMSE%, and Theil's U metrics. We can see that, based on all three comparative statistical measures, the MRJD-EGARCH (1,1) is the best model for tracking the actual time path of the WTI and Gasoline log-prices with the statistics for the MRJD-OLS being very similar. For the Heating Oil market, the best model appears to be the MRJD-OLS, which is marginally better than the MRJD-EGARCH (1,1) on the basis of the RMSE and RMSE% statistics. For all the remaining markets the model that best captures the price paths of the underlying series appears to be the MRJD-OLS, a result which is verified by all three statistical measures, with the MRJD-EGARCH exhibiting the second-best performance. It is only for the Gasoline_WTI crack spread that the MR-OLS and MR-EGARCH (1,1) models appear to perform better than the respective models incorporating jumps. Hence, our initial motivation to use Poisson jumps and to allow for two different speeds of mean reversion in the modelling procedure, in order to explain the spikier behaviour of the energy log-prices, combined with an EGARCH specification for the variance, is validated by the above findings.

Although the statistics presented above are very helpful by giving an indication on the relative quality of each model, another important criterion is how well the model captures the turning points in the data. For that, a very useful test can be a simple visual inspection of the sample price processes and the associated log-return prices (Clewlow and Strickland, 2000). Therefore, we produce a graphical comparison of the simulated prices with the actual data, plotting at first a random simulated price path and the observed data, and at second the distribution of the daily log-returns as a histogram and the daily log-returns for the average per time-step simulated prices as an overlaid line. Figure 2 shows the plot of a random simulated path for the MRJD-EGARCH (1,1) model over the actual path of the log-prices, for all energy markets. We can see that the MRJD-EGARCH (1,1) model can capture most of the major turning points in the data, tracking close enough the actual path. In particular, a major feature of our model is the fact that following a jump in the prices, the price series mean-reverts to its mean at a faster rate which is consistent with the pattern observed in the market. In addition, Figure 3 shows the distribution of the actual spot daily log-returns as a histogram and the daily log-returns for the average per time-step simulated prices as an overlaid line, for all energy markets. We observe that the MRJD-EGARCH (1,1) model captures very well the kurtosis and the skewness of the actual log-returns for almost all energy markets. This observation enhances our findings from Tables 6 and 7, where the MRJD model with an

EGARCH specification for the variance is amongst the best performing models in terms of approximating the actual returns' distribution.

6. Out-of-sample VaR evaluation

In order to evaluate the efficiency of the proposed spike models in an out-of-sample setting, we consider their performance in generating 99% one-day VaR forecasts for each one of the energy markets.[10] The period used to estimate the parametric VaR models is from 12/09/2000 to 12/09/2007 consisting of 1827 observations, whereas the period used for the out-of-sample evaluation is from 13/09/2007 to 1/02/2010 (623 observations). Based on the models that were estimated in the previous section, 100,000 Monte Carlo price paths are estimated to generate one-day ahead VaR forecasts for each one of the 623 observations in the out-of-sample period and then the one-day 99% VaR is calculated as the relevant percentile of the distribution of simulated paths. Mathematically:

$$VaR_t(\alpha) = Percentile\{r_t^s, a\} \tag{12}$$

where r_t^s represents the simulated returns at time t.

The performance of the spike models is also compared to that of two widely used benchmarks in VaR applications, the Risk Metrics and the Historical Simulation approaches. RiskMetrics (RM) uses an Exponentially Weighted Moving Average (EWMA) specification for the volatility assuming a value of $\lambda = 0.94$ for the volatility decay factor, as is widely used in the literature. The Historical Simulation (HS) method is amongst the simplest ones for estimating the VaR as it uses the past history of returns to generate the distribution of possible future returns. Under the HS methodology, the VaR with coverage rate, *a*, is calculated as the relevant percentile of the sequence of past returns, obtained non-parametrically from the data.

To select the best model in terms of its VaR forecasting power, a two stage evaluation framework is implemented. In the first stage, three statistical criteria are used to test for unconditional coverage, independence, and conditional coverage, as proposed by Christoffersen (1998). A VaR model successfully passes the first stage evaluation only when it can satisfy all three statistical tests, at the 5% or higher significance level. In the second

[10] 95% one-day VaR forecasts are also calculated but are not reported because results are very similar with the 99% forecasts that are reported in the tables.

stage, a loss function is constructed in line with Lopez (1999) and Sarma et al. (2003) to test the economic accuracy of the VaR models that have passed the first evaluation stage. The loss function is based on the notion of Expected Shortfall (ES), also termed Conditional VaR (CVaR), which measures the difference between the actual and the expected losses when a VaR violation actually occurs. Using this loss function, the models are ranked and an economic utility function able to accommodate the risk manager's needs is specified as follows:

$$LF_i = \frac{1}{T}\sum_{j=1}^{T}\left[r_j - ES_i(a)\right]^2 \tag{15}$$

$$ES_a = E[r_t|(r_t \leq -VaR_t(a))] \tag{16}$$

where the ES is defined as the average loss over the VaR violations from the N out-of-sample violations that occurred for the i_{th} VaR model, under the following conditions:

$$r_j - ES_i(a) = \begin{cases} 0, & if\ ES_i(a) \leq r_j \\ r_j - ES_i(a), & if\ r_j < ES_i(a) \end{cases} \tag{17}$$

The proposed LF uses the ES rather than the VaR measures to compare with the actual returns, as the VaR returns do not give an indication about the size of expected loss when a violation occurs. Evidence in the literature shows that ES is a more coherent risk measure than VaR (Acerbi, 2002; Inui and Kijima, 2005). The model that minimizes the total loss, hence returns the lowest LF value, is preferred relative to the remaining models. The economic evaluation framework that uses the proposed LF can provide useful information for evaluating the VaR estimates for regulatory purposes. That is because by using the ES measure in the LF, the additional information on the magnitude of a loss that exceeds the estimated VaR is incorporated into the evaluation process.

Table 8 reports the average VaR or Expected Tail Loss in percentage points, the frequency of violations or number of hits in percentage points, and the results from the second evaluation stage, i.e. the Expected Shortfall, and the Loss Function that measures the economic accuracy of the models. In the second evaluation stage only the models that pass all three of Christoffersen's tests for unconditional coverage, independence, and conditional coverage, at the 5% significance level, and thus they do not reject the null hypothesis, are indicated in bold. A 5% significance level is chosen as the acceptance threshold for the three tests, because the smaller the significance level the fewer the number of violations is, which leads to larger Type II errors that can be very costly for the risk manager. Also, the model that minimizes the total loss, hence returns the lowest LF value, is preferred relative to the remaining models, and is indicated with an asterisk. In those cases where the frequency of hits is zero the respective models are unsuitable candidates for the application of both the statistical and the economic evaluation tests. [11] A dash indicates that the test is not applicable for each respective modelling approach. In addition, in those cases where the frequency of hits is too high, above 20%, the respective models are unsuitable candidates for the application of the two statistical tests for unconditional and conditional coverage; in these cases a dash is also inserted. However, this does not mean that these models should be immediately rejected but it should be noted that consistently overestimate in the former case, and underestimate in the latter case, the actual VaR.

The results show that there is always at least one model that passes all three statistical tests at the 1% significance level; the only exception is for Natural Gas where no model is able to pass the first evaluation stage. In the majority of cases, it is only the MC simulation models that successfully pass the first evaluation stage, thus overall prevailing against the more traditional Risk Metrics and Historical Simulation methodologies. For WTI and Heating Oil, it is the MRJD-GARCH model that passes the first evaluation stage and also delivers in the second stage the lowest loss function value. In addition, for the remaining energy markets, i.e. Gasoline, Propane, and the two crack spreads with WTI, it is the MRJD-EGARCH model that outperforms all competing models; the only exception is for PJM where the Historical Simulation method is the best performing one. Therefore, whenever a risk manager wants to choose a single approach for calculating the VaR for all energy commodities that he/ she holds, as it is usually the case in practice, the results indicate that the MC simulations incorporating jumps and a GARCH or an EGARCH volatility specifications, as proposed in this paper, are the most reasonable, efficient, and consistent candidates.

[11] The mean reverting models without jumps are not included in the analysis as they do not provide any hits during the first statistical evaluation stage.

7. Conclusions

In this paper we examine the behaviour of spot prices in the eight energy markets that trade futures contracts on NYMEX. Given the stylised properties of those markets, we propose a mean-reverting spike model that incorporates two different speeds of mean reversion to capture the fast mean-reverting behaviour of prices after a jump occurs and the slower mean reversion rate of the diffusive part of the model. We also extend this model to incorporate time-varying volatility in its specification, modelled as an EGARCH process. Estimation results indicate the presence of a "leverage effect" for WTI, Heating Oil, and Heating Oil – WTI crack spread spot log-price returns, whereas for the remaining energy markets the presence of an "inverse leverage" effect is found.

The comparison of the different models used in this paper is done using 100,000 Monte Carlo simulations in each case. Results indicate that the inclusion of Poisson jumps to the mean reverting model, in combination with the use of a different speed of mean reversion after a jump occurs for a duration equal to the half-life of the jumps' returns, improves the fit significantly for all energy markets. Our modelling approach captures very well both the skewness and kurtosis of the actual series. Furthermore, the addition of the EGARCH (1,1) specification for the variance improves significantly the fit of the simulated returns to the actual distributions, for most of the energy markets under investigation. This finding is validated by the reported Kolmogorov-Smirnov statistics, as well as by comparing visually the simulated to the actual price series. Moreover, the proposed models, incorporating jumps and a GARCH or an EGARCH volatility specifications, are the most efficient and consistent candidates for estimating VaR for the majority of the energy markets examined in this paper. Hence, overall, our modelling approach for energy pricing combined with the findings of this paper is relevant for both policymakers and market participants as it can be applied for forecasting, risk management, derivatives pricing, and policy development and monitoring purposes.

References

Acerbi, C., 2002. Spectral measures of risk: a coherent representation of subjective risk aversion, Journal of Banking and Finance, 26 (7), 1505-1518.

Askari, H., Krichene, N., 2008. Oil price dynamics (2002-2006). Energy Economics 30; 2134-2153.

Baker, M. P., Mayfield, E. S., Parsons, J. E., 1998. Alternative models of uncertain commodity prices for use with modern asset pricing. Energy Journal 19 (1); 115-148.

Ball, C.A., Torous, W.N., 1983. A Simplified Jump Process for Common Stock Returns. Journal of Financial and Quantitative Analysis 18 (1); 53-65.

Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in Deutsche mark options. Review of Financial Studies 9 (1); 69-107.

Baumeister, C., Peersman, G., 2008. Time-varying effects of oil supply shocks on the US economy. Working Paper; Universiteit Gent.

Bessembinder, H., Coughenour, J.F., Seguin, P.J., Smoller, M.M., 1995. Mean Reversion in Equilibrium Asset Prices: Evidence from the Futures Term Structure. Journal of Finance 50 (1); 361-375.

Bierbrauer, M., Menn, C, Rachev, S.T., Truck, S., 2007. Spot and derivative pricing in the EEX power market. Journal of Banking and Finance 31; 3462-3485.

Black, F., Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81; 637-654.

Black, F., 1976. Studies in stock price volatility changes. Proceedings of the business and economics statistics section. American Statistical Association; 177-181.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31; 307–327.

Brennan, M. J., 1991. The price of convenience yield and the valuation of commodity contingent claims. In: D. Lund and B. Oksendal (Eds.), Stochastic models and option values, New York: Elsevier Science Publishers B. V., North Holland; 1991. pp. 33-71.

Cartea, A., Figueroa, M., 2005. Pricing in electricity markets: a mean reverting jump diffusion model with seasonality. Applied Mathematical Finance 12 (4); 313-335.

Cartea, A., Villaplana, P., 2008. Spot price modeling and the valuation of electricity forward contracts: The role of demand and capacity. Journal of Banking and Finance, 32; 2502-2519.

Christoffersen, Peter F., 1998. Evaluating Interval Forecasts. International Economic Review, 39, 841-862.

Clewlow, L., Strickland, C. Energy Derivatives Pricing and Risk Management. Lacima Publications (Published in association with Enron Corp.); 2000.

Clewlow, L., Strickland, C., Kaminski, V., 2000. Spot simulation processing. Energy & Power Risk Management, Risk Waters Group.

Clewlow, L., Strickland, C., Kaminski, V., 2001. Extending Mean Reversion Jump Diffusion. Energy & Power Risk Management, Risk Waters Group.

Cortazar, G., Schwartz, E., 1994. The Valuation of Commodity Contingent Claims. The Journal of Derivatives 1; 27-39.

Dias, M.A.G., 2003. Investment in information in petroleum, real options and revelation. Working Paper; Pontifícia Universidade Católica do Rio de Janeiro.

Dickey, D.A., W.A. Fuller, 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association 74; 427–431.

Dixit, A. K., Pindyck, R. S. Investment under uncertainty. Princeton University Press; 1994.

Engle, R. F., 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. Econometrica 50; 987-1008.

Engle, R. F., Patton, A. J., 2001. What good is a volatility model? Quantitative Finance 1; 237-245.

Eydeland, A., Wolyniec, K. Energy and power risk management: new developments in modeling, pricing, and hedging. John Wiley & Sons, Inc.; Hoboken; 2003.

Geman, H. Commodities and commodity derivatives: modeling and pricing for agricultural, metal and energy. Wiley Finance; 2005.

Geman, H., Roncoroni, A., 2006. Understanding the fine structure of electricity prices. The Journal of Business 79 (3); 1225-1262.

Gibson, R., Schwartz, E., 1990. Pricing of oil contingent claims. Journal of Finance 45; 959-976.

Inui, K. and Kijima, M., 2005. On the significance of Expected Shortfall as a coherent risk measure, Journal of Banking and Finance, 29, 853-864.

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters **6** (3); 255–259; doi:10.1016/0165-1765(80)90024-5.

JP Morgan, 1996. RiskMetrics, Technical Document, New York.

Kanamura, T., 2009. A supply and demand based volatility model for energy prices. Energy Economics; doi: 10.1016/j.eneco.2009.04.001.

Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Economics 27; 791-817.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of Econometrics 54; 159–178.

Lopez, J.A., 1999. Regulatory evaluation of Value-at-Risk models. Journal of Risk, 1, 37-64.

Lucia, J.J., Schwartz, E.S., 2002. Electricity prices and power derivatives: evidence from the Nordic power exchange. Review of Derivatives Research 5; 5-50.

Ljung, G.M., Box, G.E.P., 1978. On a Measure of a Lack of Fit in Time Series Models. Biometrika 65; 297-303.

Merton, R.C., 1973. The Theory of Rational Option Pricing. Bell Journal of Economics and Management Science 4; 141-183.

Merton, R.C., 1976. Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics 3; 125-144.

Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica 59; 347-370.

Nomikos, N., Soldatos, O., 2008. Using affine jump diffusion models for modeling and pricing electricity derivatives. Applied Mathematical Finance 15 (1); 41-71.

Phillips, P., Perron, P., 1988. Testing for a unit root in time series regressions. Biometrika 75; 335-346.

Pindyck, R. S., 1999. The Long-Run Evolution of Energy Prices. The Energy Journal 20 (2).

Pindyck, R. S., Rubinfeld, D.L. Econometric Models and Economic Forecasts. 4^{th} ed., Irwin/McGraw-Hill; 1998.

Pilipovic, D. Energy risk: Valuing and managing energy derivatives. McGraw-Hill; 1998.

Sarma, M., Thomas, S., Shah, A., 2003. Selection of Value-at-Risk models. Journal of Forecasting 22, 337-358.

Schwartz E.S., 1997. The Stochastic Behaviour of Commodity Prices: Implications for Valuation and Hedging. Journal of Finance LII (3); 923-973.

Theil, H. Economic Forecasts and Policy. North-Holland: Amsterdam; 1961.

Weron, R. and Misiorek, A., 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International Journal of Forecasting 24 (4), 744-763.

Table 1: Empirical models of energy prices

"GBM" stands for Geometric Brownian Motion; "MR" for Mean Reversion; "MRJD" for Mean Reversion Jump Diffusion, "OLS" for Ordinary Least Squares (constant volatility)

1	GBM
2	MR-OLS
3	MR-GARCH (1,1)
4	MR-EGARCH (1,1)
5	MRJD-OLS
6	MRJD-GARCH (1,1)
7	MRJD-EGARCH (1,1)

Table 2: Descriptive statistics of energy markets

Descriptive statistics and the properties of the logarithmic spot prices and their first differences (returns) are presented in Panels A and B, respectively. *, **, *** denote significance at the 10%, 5% and 1% significance level, respectively. Two tests for unit root non-stationarity, the Augmented Dickey-Fuller (ADF; Dickey and Fuller, 1979) and the Philips-Perron (PP; Phillips and Perron, 1988), and one test for stationarity, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS; Kwiatkowski et. al, 1992), are employed. The Jarque-Bera (1980) test for normality on the logarithmic differences is χ^2 distributed with 2 degrees of freedom. Q(k) is the Ljung-Box (1978) Q-statistic test for k_{th} order autocorrelation. The $Q^2(k)$-statistic is the Engle's (1982) ARCH test. Both tests are χ^2 distributed with k degrees of freedom. Daily data from 12/9/2000 to 12/9/2007.

Panel A: Logarithmic levels

	WTI	HO	GASOLINE	CS_GASOLINE_WTI	CS_HO_WTI	NG	PROPANE	PJM
Mean Spot Level ($)	$40.01	$46.15	$48.63	$109.17	$106.94	$5.44	$67.58	$49.81
Mean (μ)	3.6890	3.8320	3.8843	4.6929	4.6723	1.6939	4.2133	3.9081
Mean (excl. jumps)	3.6892	3.8318	3.8845	4.6927	4.6722	1.6932	4.2131	3.9067
Maximum	4.381	4.511	4.793	5.019	4.815	2.944	4.851	5.701
Minimum	2.861	2.981	3.011	4.568	4.604	0.528	3.287	3.002
Standard Deviation	0.394	0.415	0.395	0.062	0.041	0.401	0.383	0.399
Skewness	0.054	0.039	0.110	1.685	0.501	-0.377	-0.321	0.198
Kurtosis	1.691	1.620	2.073	6.259	2.383	3.378	2.106	3.291
KPSS	4.917	4.730	4.621	1.269	3.102	2.650	4.297	2.963
ADF	-0.657	-0.711	-1.384	-4.581***	-3.575***	-2.370	-0.795	-4.853***
	(0.855)	(0.842)	(0.591)	(0.000)	(0.006)	(0.151)	(0.820)	(0.000)
PP	-0.441	-0.668	-1.371	-4.488***	-3.505***	-0.789	-2.452	-8.644***
	(0.900)	(0.853)	(0.598)	(0.000)	(0.008)	(0.821)	(0.128)	(0.000)

Panel B: Logarithmic differences (returns)

	WTI	HO	GASOLINE	CS_GASOLINE_WTI	CS_HO_WTI	NG	PROPANE	PJM
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.106	0.119	0.183	0.118	0.057	0.623	0.363	0.963
Minimum	-0.172	-0.188	-0.178	-0.165	-0.053	-0.570	-0.244	-1.428
Standard Deviation	0.023	0.026	0.030	0.013	0.008	0.049	0.024	0.152
Annualised Volatility	0.365	0.410	0.480	0.210	0.130	0.778	0.376	2.410
Skewness	-0.454	-0.269	-0.264	-0.920	0.374	0.732	1.609	0.059
Kurtosis	6.485	6.695	6.759	29.302	10.674	32.845	45.154	12.780
Jarque-Berra	987.237	1061.348	1096.627	52893.450	4523.756	67970.450	135982.000	7281.879
KPSS	0.162	0.165	0.061	0.021	0.046	0.035	0.060	0.086
ADF	-43.5919***	-45.7299***	-43.3930***	-45.279***	-34.019***	-25.7665***	-44.06193***	-25.5479***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
PP	-43.9383***	-45.8342***	-43.4004***	-45.300***	-50.285***	-41.2204***	-44.04424***	-104.2735***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Q (1)	0.786	8.552***	0.427	6.367**	15.231***	2.564	1.7917	11.836***
Q (20)	22.502	29.155*	19.056	49.511***	70.225***	116.930***	38.154***	199.990***
Q^2 (1)	13.763***	46.025***	43.495***	197.110***	59.978***	438.58***	365.75***	14.014***
Q^2 (20)	55.904***	154.29***	171.08***	320.130***	280.530***	935.5***	390.28***	217.01***

Table 3: Estimated jump parameters, mean reversion rates, volatility, and half-lives

The filtered series exclude all returns that have been identified as jumps (more than three times the standard deviation of the smooth returns). Φ is the daily frequency of a jump occurring, σ_J is the daily standard deviation of jump returns, and $\bar{\kappa}_J$ the average size of jump returns. The diffusive mean reversion rate α, is estimated using eq. 6 after running the regression of eq. 5. The mean reversion rate used after a jump has occurred α_{JD}, for a period of time equal to the half-life of jump returns, is estimated using eq. 9 after running the regression of eq. 8. Also, σ is the daily standard deviation of log-price differences, as estimated from eq. 7 for the un-filtered and filtered series, respectively. All estimates for the half-lives of both the smooth and jumpy returns are calculated using eq. 10. The half-lives of the jumpy returns, in days, are the respective durations we are using in our MRJD models for the higher mean reversion rate (α_{JD}) after a jump occurs.

Panel A: Jump parameters used in the MRJD models

	Φ_{daily}	σ_J	$\bar{\kappa}_J$
WTI	0.0192	0.0725	-0.0460
HO	0.0159	0.0899	0.0086
GASOLINE	0.0235	0.1078	-0.0089
CS_GASOLINE_WTI	0.1873	0.0305	0.0208
CS_HO_WTI	0.0405	0.0277	0.0065
NG	0.0581	0.1614	0.0627
PROPANE	0.0476	0.0816	0.0176
PJM	0.0728	0.5194	-0.0214

Panel B: Mean reversion rates, daily st. deviations, and half-lives of smooth and jumpy returns

	Un-filtered series (MR)	Filtered Series (MRJD)	Half-lives for MRJD models, in days
WTI			
α	0.001	0.001	998
α_{JD}	-	0.019	36
σ	0.023	0.022	
HO			
α	0.001	0.001	771
α_{JD}	-	0.010	67
σ	0.026	0.024	
GASOLINE			
α	0.002	0.002	362
α_{JD}	-	0.021	34
σ	0.030	0.027	
CS_GASOLINE_WTI			
α	0.023	0.012	60
α_{JD}	-	0.026	26
σ	0.013	0.009	
CS_HO_WTI			
α	0.020	0.013	55
α_{JD}	-	0.041	17
σ	0.008	0.007	
NG			
α	0.007	0.004	155
α_{JD}	-	0.010	72
σ	0.049	0.038	
PROPANE			
α	0.001	0.000	2635
α_{JD}	-	0.008	87
σ	0.024	0.017	
PJM			
α	0.075	0.055	13
α_{JD}	-	0.115	6
σ	0.158	0.132	

Table 4: Comparison of the arrival rate of the actual daily jumps' series to the arrival rate of a Poisson generated series.

Comparison of the actual distribution of daily jumps as identified by the R-F methodology, to the distribution of a series of jumps generated by a Poisson process with a frequency equal to the reported frequency of jump occurrence in panel A of Table 3, for each energy market. The null hypothesis of the two-sample Kolmogorov-Smirnov (K-S) test is that the two samples are from the same continuous distribution, at the 5% significance level.

	K-S
WTI	0.0011
HO	0.0022
GASOLINE	0.0005
CS_GASOLINE_WTI	0.0011
CS_HO_WTI	0.0044
NG	0.0038
PROPANE	0.0006
PJM	0.0033

Table 5: GARCH and EGARCH coefficient estimates from the un-filtered series

The regression results of equation 5 are presented, considering a GARCH and an EGARCH estimate for the variance, respectively. The regression is applied to both the un-filtered and filtered series, with the estimates used for the MR and MRJD models, respectively. Results are qualitatively similar and only those estimated from the un-filtered historical series are reported in the table. p-values are in brackets. The GARCH and EGARCH volatility equations are the following:

$$\sigma_t = \sqrt{\beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 * \sigma_{t-1}^2} \quad [GARCH(1,1)]$$

$$\sigma_t = \sqrt{e^{\beta_0 + \beta_1 * \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} + \beta_2 * \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_3 * \ln\left(\sigma_{t-1}^2\right)}} \quad [EGARCH(1,1)]$$

	WTI	**HO**	**GASOLINE**	**CS_GASOLINE_WTI**	**CS_HO_WTI**	**NG**	**PROPANE**	**PJM**
GARCH(1,1)								
β_0	0.00003	0.00006	0.00012	0.00000	0.00000	0.00006	0.00004	0.00066
	(0.00001)	(0.00001)	(0.00002)	(0.00000)	(0.00000)	(0.00001)	(0.00000)	(0.00011)
β_1	0.05992	0.09713	0.09090	0.13803	0.15535	0.13596	0.14783	0.13640
	(0.00695)	(0.01123)	(0.01055)	(0.00609)	(0.01002)	(0.01522)	(0.00668)	(0.01310)
β_2	0.88950	0.81687	0.78194	0.88450	0.84440	0.86011	0.77278	0.84627
	(0.01741)	(0.02545)	(0.03025)	(0.00342)	(0.00812)	(0.01312)	(0.00779)	(0.01301)
EGARCH(1,1)								
β_0	-0.69575	-0.71312	-0.86639	-0.31057	-1.34905	-0.32579	-1.58968	-0.30804
	(0.10537)	(0.11634)	(0.15644)	(0.02492)	(0.12466)	(0.04098)	(0.09414)	(0.02651)
β_1	0.10618	0.19570	0.19953	0.20868	0.35897	0.21273	0.36064	0.24126
	(0.01843)	(0.01651)	(0.02042)	(0.01089)	(0.02101)	(0.01886)	(0.01007)	(0.01734)
β_2	-0.10648	-0.00630	0.00658	0.06972	-0.03414	0.07314	0.02848	0.03709
	(0.01404)	(0.01002)	(0.01179)	(0.00822)	(0.01211)	(0.00896)	(0.00887)	(0.01212)
β_3	0.91928	0.92322	0.89790	0.98322	0.88680	0.97227	0.82680	0.96604
	(0.01341)	(0.01486)	(0.02098)	(0.00256)	(0.01208)	(0.00543)	(0.01207)	(0.00484)

Table 6: Distributional comparison of the actual spot log-price returns to the average per time-step simulated path

Distributional comparison of the actual spot logarithmic-price returns to the average per time-step simulated path for each model specification, $S_t^s = \sum_{\omega=1}^{n} \frac{S_{t,\omega}^s}{n}$, where $S_{t,\omega}$ is the simulated spot price of path ω at time t, and n is the number of MC simulations. K-S is the Kolmogorov-Smirnov (K-S) two-sample test statistic; an asterisk (*) indicates that we accept the null that the two samples are from the same continuous distribution, at the 5% significance level. The models with the smallest K-S test-statistic value are indicated with a (+).

	Mean	Median	Max	Min	Std. Dev.	Skewness	Kurtosis	K-S
WTI								
Actual Path	0.00	0.00	0.11	-0.17	0.02	-0.45	6.48	
GBM	0.00	0.00	0.08	-0.08	0.02	0.00	3.00	0.501
MR-OLS	0.00	0.00	0.12	-0.17	0.03	-0.16	3.87	0.058
MR-GARCH(1,1)	0.00	0.00	0.12	-0.17	0.03	-0.17	3.94	0.059
MR-EGARCH (1,1)	0.00	0.00	0.13	-0.17	0.03	-0.16	3.94	0.055
MRJD-OLS	0.00	0.00	0.13	-0.21	0.03	-0.45	5.39	0.056
MRJD-GARCH(1,1)	0.00	0.00	0.14	-0.21	0.04	-0.34	4.91	0.055
MRJD-EGARCH (1,1)	0.00	0.00	0.13	-0.21	0.03	-0.42	5.28	0.054+
HO								
Actual Path	0.00	0.00	0.12	-0.19	0.03	-0.27	6.69	
GBM	0.00	0.00	0.09	-0.09	0.03	0.00	3.00	0.491
MR-OLS	0.00	0.00	0.14	-0.19	0.04	-0.10	3.92	0.067
MR-GARCH(1,1)	0.00	0.00	0.14	-0.19	0.04	-0.10	4.03	0.065
MR-EGARCH (1,1)	0.00	0.00	0.14	-0.20	0.04	-0.09	3.92	0.065
MRJD-OLS	0.00	0.00	0.20	-0.19	0.04	0.08	4.97	0.060
MRJD-GARCH(1,1)	0.00	0.00	0.21	-0.19	0.04	0.07	4.76	0.057
MRJD-EGARCH (1,1)	0.00	0.00	0.20	-0.19	0.04	0.08	4.83	0.056+
GASOLINE								
Actual Path	0.00	0.00	0.18	-0.18	0.03	-0.26	6.76	
GBM	0.00	0.00	0.10	-0.10	0.03	0.00	3.00	0.484
MR-OLS	0.00	0.00	0.20	-0.20	0.04	-0.09	3.94	0.044*
MR-GARCH(1,1)	0.00	0.00	0.20	-0.19	0.04	-0.09	4.00	0.045
MR-EGARCH (1,1)	0.00	0.00	0.20	-0.20	0.04	-0.09	3.91	0.046
MRJD-OLS	0.00	0.00	0.24	-0.26	0.04	-0.11	6.04	0.044*+
MRJD-GARCH(1,1)	0.00	0.00	0.24	-0.26	0.04	-0.10	5.79	0.045
MRJD-EGARCH (1,1)	0.00	0.00	0.24	-0.26	0.04	-0.10	5.85	0.045
CS_GASOLINE_WTI								
Actual Path	0.00	0.00	0.12	-0.16	0.01	-0.92	29.30	
GBM	0.00	0.00	0.05	-0.05	0.01	0.00	3.00	0.463
MR-OLS	0.00	0.00	0.12	-0.16	0.02	-0.32	9.47	0.029*
MR-GARCH(1,1)	-0.01	0.00	2.34	-2.57	0.48	-0.27	8.62	0.297
MR-EGARCH (1,1)	0.00	0.00	0.12	-0.16	0.02	-0.29	8.52	0.025*
MRJD-OLS	0.00	0.00	0.11	-0.08	0.02	1.15	6.69	0.023*+
MRJD-GARCH(1,1)	0.00	0.00	0.42	-0.43	0.08	-0.03	6.70	0.028*
MRJD-EGARCH (1,1)	0.00	0.00	0.13	-0.11	0.03	0.44	4.58	0.024*
CS_HO_WTI								
Actual Path	0.00	0.00	0.06	-0.05	0.01	0.37	10.67	
GBM	0.00	0.00	0.03	-0.03	0.01	0.00	3.00	0.474
MR-OLS	0.00	0.00	0.06	-0.06	0.01	0.13	4.89	0.031*
MR-GARCH(1,1)	0.00	0.00	0.11	-0.12	0.02	0.02	8.23	0.031*
MR-EGARCH (1,1)	0.00	0.00	0.06	-0.06	0.01	0.10	4.66	0.032*
MRJD-OLS	0.00	0.00	0.08	-0.06	0.01	0.45	8.28	0.029*
MRJD-GARCH(1,1)	0.00	0.00	0.13	-0.13	0.02	0.03	7.06	0.023*
MRJD-EGARCH (1,1)	0.00	0.00	0.08	-0.07	0.01	0.20	6.33	0.022*+

Table 6 cont.

NG								
Actual Path	0.00	0.00	0.62	-0.57	0.05	0.73	32.85	
GBM	0.00	0.00	0.17	-0.17	0.05	0.00	3.00	0.453
MR-OLS	0.00	0.00	0.62	-0.57	0.07	0.26	10.42	0.059
MR-GARCH(1,1)	0.00	0.00	0.73	-0.73	0.11	-0.06	8.83	0.070
MR-EGARCH (1,1)	0.00	0.00	0.63	-0.57	0.08	0.13	7.44	0.056
MRJD-OLS	0.00	0.00	0.49	-0.37	0.07	0.90	9.59	0.049
MRJD-GARCH(1,1)	0.00	0.00	0.58	-0.55	0.11	0.12	5.66	0.052
MRJD-EGARCH (1,1)	0.00	0.00	0.51	-0.40	0.08	0.43	6.16	0.049^{+}
PROPANE								
Actual Path	0.00	0.00	0.36	-0.24	0.02	1.61	45.15	
GBM	0.00	0.00	0.08	-0.08	0.02	0.00	3.00	0.505
MR-OLS	0.00	0.00	0.36	-0.24	0.03	0.57	13.52	0.110
MR-GARCH(1,1)	0.00	0.00	0.36	-0.24	0.03	0.60	14.71	0.107
MR-EGARCH (1,1)	0.00	0.00	0.36	-0.25	0.03	0.55	13.18	0.108
MRJD-OLS	0.00	0.00	0.23	-0.19	0.03	0.55	10.39	0.096
MRJD-GARCH(1,1)	0.00	0.00	0.24	-0.21	0.04	0.20	6.34	0.093
MRJD-EGARCH (1,1)	0.00	0.00	0.23	-0.20	0.04	0.33	7.19	0.092^{+}
PJM								
Actual Path	0.00	0.00	0.96	-1.43	0.15	0.06	12.78	
GBM	0.00	0.00	0.52	-0.52	0.15	0.00	2.99	0.467
MR-OLS	0.00	0.00	1.09	-1.43	0.22	0.02	5.34	0.044*
MR-GARCH(1,1)	0.00	0.00	1.16	-1.46	0.23	-0.01	5.97	0.041*
MR-EGARCH (1,1)	0.00	0.00	1.14	-1.43	0.24	-0.02	4.99	0.039^{*+}
MRJD-OLS	0.00	0.00	1.42	-1.46	0.23	-0.06	8.03	0.046
MRJD-GARCH(1,1)	0.00	0.00	1.63	-1.69	0.31	-0.09	6.00	0.041*
MRJD-EGARCH (1,1)	0.00	0.00	1.46	-1.51	0.26	-0.11	6.57	0.043*

Table 7: Comparison of the models' goodness of fit to the actual spot log-prices

Simulation error statistics on the difference between actual versus average simulated price paths. RMSE, RMSE %, and Theil's U are respectively calculated as:

$$RMSE=\sqrt{\frac{\sum_{t=1}^{T}\left(S_t^s-S_t^a\right)^2}{T}}\ ,\ RMSE\%=\sqrt{\frac{\sum_{t=1}^{T}\left(\frac{S_t^s-S_t^a}{S_t^a}\right)^2}{T}}\quad\text{and}\quad U=\sqrt{\frac{\sum_{t=1}^{T}\left(S_t^s-S_t^a\right)^2}{T}}\Bigg/\sqrt{\frac{\sum_{t=1}^{T}\left(S_t^s\right)^2}{T}}+\sqrt{\frac{\sum_{t=1}^{T}\left(S_t^a\right)^2}{T}}$$

where $S_t^s=\sum_{\omega=1}^{n}\frac{S_{t,\omega}^s}{n}$ is the average of the simulated values at time t across all possible paths, $S_{t,\omega}$ is the simulated spot price of path ω at time t, *n* is the number of MC simulations, S_t^a is the actual value on any given time-step, and T is the number of discretised periods in the simulation.

	GBM	MR-OLS	MR-GARCH(1,1)	MR-EGARCH (1,1)	MRJD-OLS	MRJD-GARCH(1,1)	MRJD-EGARCH (1,1)
WTI							
RMSE	0.695	0.652	0.651	0.652	0.385	0.399	0.387
RMSE %	0.188	0.169	0.168	0.168	0.099	0.103	0.100
Theil's U	0.090	0.096	0.095	0.096	0.053	0.055	0.053
HO							
RMSE	0.792	0.634	0.635	0.666	0.379	0.390	0.385
RMSE %	0.207	0.158	0.159	0.167	0.096	0.099	0.098
Theil's U	0.098	0.088	0.089	0.093	0.050	0.051	0.051
GASOLINE							
RMSE	0.860	0.528	0.524	0.554	0.377	0.382	0.380
RMSE %	0.218	0.131	0.130	0.138	0.095	0.096	0.096
Theil's U	0.108	0.071	0.070	0.074	0.049	0.050	0.049
CS_GASOLINE_WTI							
RMSE	0.361	0.067	7.629	0.075	0.166	0.401	0.177
RMSE %	0.077	0.014	1.620	0.016	0.035	0.085	0.038
Theil's U	0.039	0.007	0.668	0.008	0.017	0.043	0.019
CS_HO_WTI							
RMSE	0.224	0.048	0.092	0.054	0.045	0.105	0.054
RMSE %	0.048	0.010	0.020	0.012	0.010	0.022	0.012
Theil's U	0.024	0.005	0.010	0.006	0.005	0.011	0.006
NG							
RMSE	1.371	0.477	1.263	0.627	0.508	0.814	0.566
RMSE %	0.857	0.301	0.781	0.392	0.376	0.539	0.397
Theil's U	0.377	0.145	0.324	0.195	0.135	0.233	0.155
PROPANE							
RMSE	0.739	0.573	0.558	0.590	0.327	0.386	0.353
RMSE %	0.178	0.131	0.128	0.135	0.080	0.092	0.085
Theil's U	0.084	0.072	0.070	0.074	0.038	0.046	0.042
PJM							
RMSE	4.051	0.497	0.565	0.593	0.546	0.930	0.641
RMSE %	1.019	0.126	0.144	0.151	0.140	0.238	0.164
Theil's U	0.449	0.064	0.074	0.078	0.071	0.124	0.084

Table 8: 99% VaR results for long positions

99% VaR results for long positions across all energy commodities. The table reports the average VaR or Expected Tail Loss (ETL) in percentage points, the frequency of violations or number of hits in percentage points, and the results from the second evaluation stage, i.e. the Expected Shortfall (ES), and the Loss Function (LF) that measures the economic accuracy of the models. In the second evaluation stage only the models that pass all three of Christoffersen's tests for unconditional coverage, independence, and conditional coverage, at the 5% significance level, are indicated in bold. Also, the model that minimizes the Loss Function, is preferred relative to the remaining models, and is indicated with an asterisk. In those cases where the frequency of hits is zero the respective models are unsuitable candidates for the application of both statistical and economic evaluation tests. A dash indicates that the test is not applicable for the respective modelling approach.

		RM	HS	MCS-GBM	MCS-MRJD-OLS	MCS-MRJD-GARCH	MCS-MRJD-EGARCH
WTI	Avg VaR (ETL)	2.30%	2.03%	2.62%	2.78%	2.27%	2.45%
	No Hits (%)	31.46%	4.17%	5.14%	1.61%	1.12%	1.44%
	ES	-3.30%	-8.33%	-7.92%	-5.90%	-6.44%	-6.37%
	LF (x10^4)	1.373	0.090	0.120	**0.402**	**0.298***	**0.311**
HO	Avg VaR (ETL)	2.04%	2.10%	2.14%	2.36%	1.93%	1.81%
	No Hits (%)	32.10%	2.57%	2.89%	1.12%	0.80%	1.12%
	ES	-2.82%	-8.30%	-8.10%	-8.06%	-8.61%	-8.06%
	LF (x10^4)	1.059	0.041	0.048	**0.050**	**0.031***	**0.050**
GASOLINE	Avg VaR (ETL)	2.45%	2.37%	2.64%	3.80%	3.97%	4.06%
	No Hits (%)	30.50%	1.93%	3.53%	0.96%	0.80%	0.80%
	ES	-3.47%	11.00%	-9.63%	-10.33%	-12.64%	-12.64%
	LF (x10^4)	1.808	0.136	0.207	**0.167**	**0.088**	**0.088***
CS-GASOLINE-WTI	Avg VaR (ETL)	1.27%	2.41%	2.38%	2.73%	-	2.93%
	No Hits (%)	31.30%	2.73%	4.98%	2.25%	0.00%	0.80%
	ES	-1.80%	-6.72%	-5.45%	-5.97%	-	-9.83%
	LF (x10^4)	1.000	0.146	0.232	0.192	-	**0.033***
CS-HO-WTI	Avg VaR (ETL)	0.87%	1.80%	1.57%	2.22%	1.67%	2.41%
	No Hits (%)	31.46%	2.73%	5.14%	2.57%	0.32%	1.44%
	ES	-1.25%	-4.56%	-3.49%	-4.36%	-7.64%	-5.50%
	LF (x10^4)	0.420	0.064	0.120	0.073	0.004	**0.034***
NG	Avg VaR (ETL)	2.84%	4.17%	4.49%	3.11%	-	-
	No Hits (%)	32.91%	0.80%	0.80%	0.32%	0.00%	0.00%
	ES	-3.97%	15.88%	-15.88%	-18.51%	-	-
	LF (x10^4)	2.327	0.108	0.108	0.050	-	-
PROPANE	Avg VaR (ETL)	1.87%	4.13%	3.91%	3.72%	3.74%	4.94%
	No Hits (%)	29.53%	1.77%	2.57%	1.28%	0.48%	0.64%
	ES	-1.54%	-4.59%	-4.27%	-2.01%	-2.13%	-2.43%
	LF (x10^4)	2.590	1.114	1.204	**2.239**	**2.157**	**1.971***
PJM	Avg VaR (ETL)	10.44%	13.42%	14.19%	-	-	-
	No Hits (%)	25.84%	0.80%	1.61%	0.00%	0.00%	0.00%
	ES	13.80%	58.75%	-49.38%	-	-	-
	LF (x10^4)	30.366	**0.739***	**1.858**	-	-	-

Figure 1: Graphs of daily log-spot and first log-differences for the crude oil, gasoline oil, and heating oil (WTI, Gasoline, HO), the two 1-1 crack spreads with the crude oil (CS_Gasoline_WTI, CS_HO_WTI), and for the electricity, natural gas, and propane markets (PJM, NG, Propane). Data period is from 12/09/2000 to 12/09/2007.

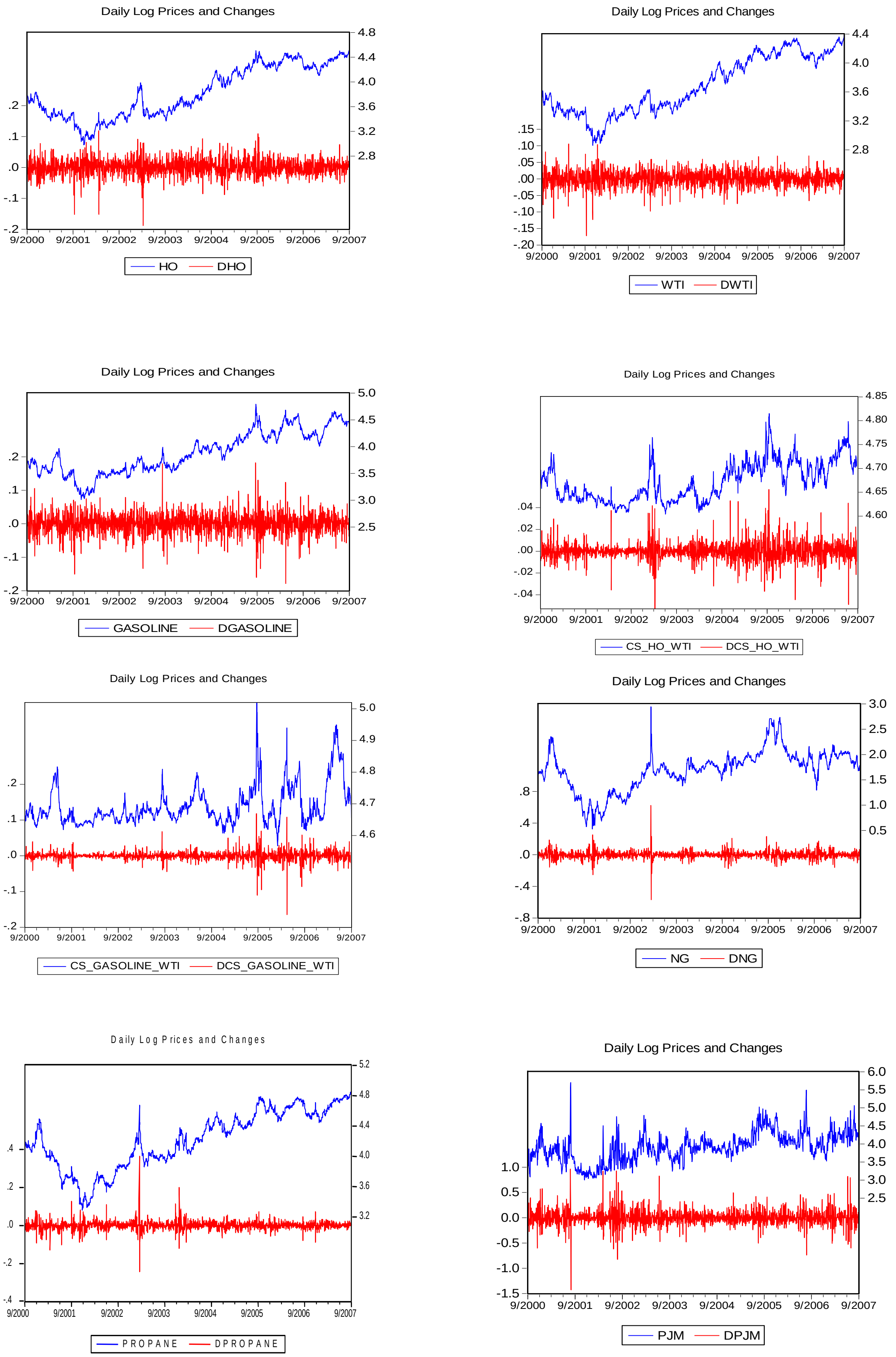

Figure 2: Random simulated path of spot log-prices from the MRJD-EGARCH (1,1) model plotted against the actual path, for all energy markets.

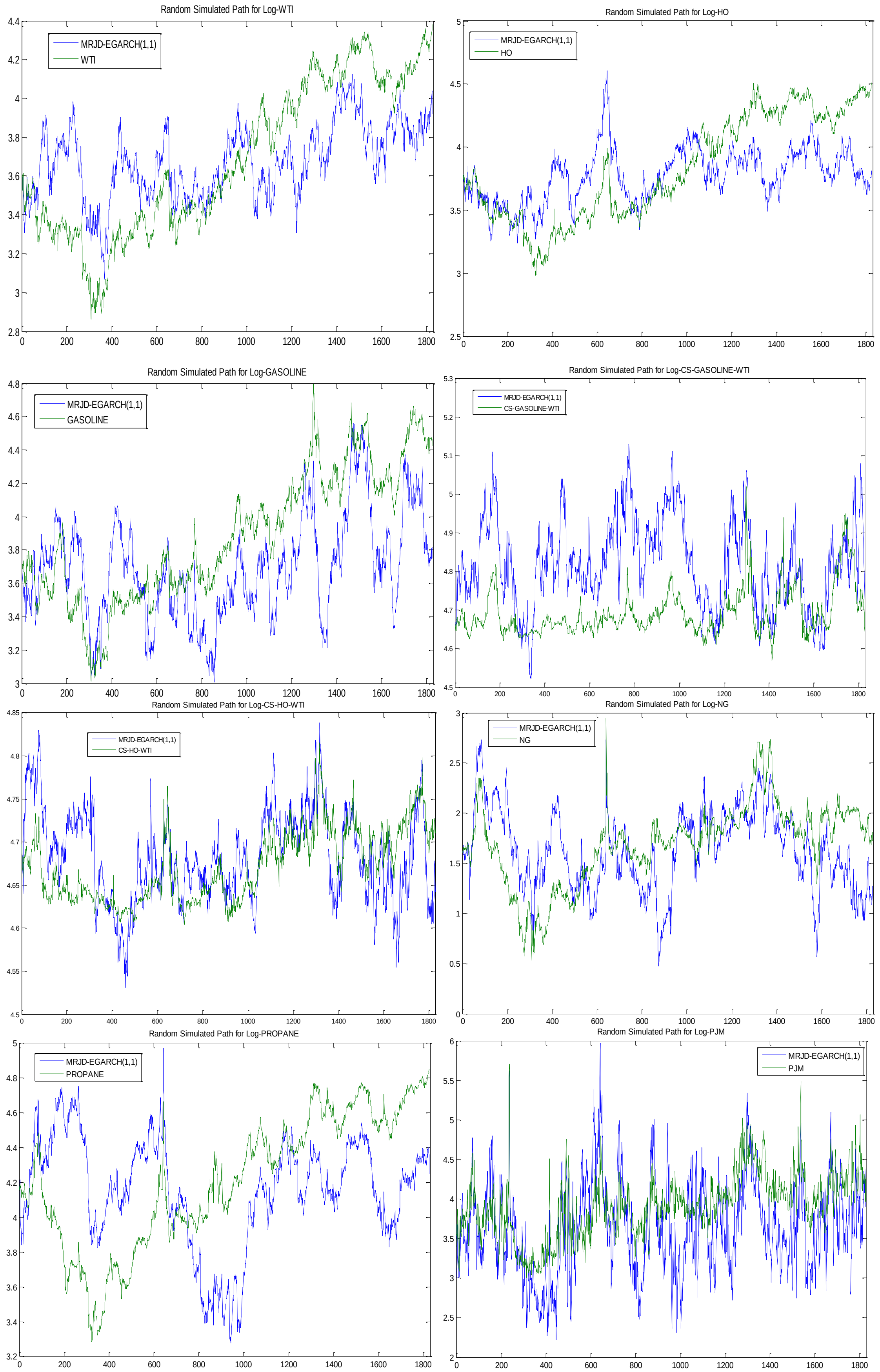

Figure 3: Histogram of the average simulated spot log-price returns per time-step for the MRJD-EGARCH (1,1) model plotted as a solid line against the actual returns, for all energy markets.